Environmental Science and Ecotechnology 20 (2024) 100353

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

journal homepage: www.journals.elsevier.com/environmental-science-and-ecotechnology/

Original Research

Integrated genomics provides insights into the evolution of the polyphosphate accumulation trait of *Ca.* Accumulibacter

Xiaojing Xie [a, 1], Xuhan Deng [a, 1], Liping Chen [a], Jing Yuan [a], Hang Chen [a], Chaohai Wei [a, c], Xianghui Liu [b, c], Stefan Wuertz [b, c, **], Guanglei Qiu [a, b, d, e, *]

[a] School of Environment and Energy, South China University of Technology, Guangzhou, 510006, China
[b] Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, 637551, Singapore
[c] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, 639798, Singapore
[d] Guangdong Provincial Key Laboratory of Solid Wastes Pollution Control and Recycling, Guangzhou, 510006, China
[e] The Key Lab of Pollution Control and Ecosystem Restoration in Industry Clusters, Ministry of Education, Guangzhou, 510006, China

ARTICLE INFO

Article history:
Received 14 March 2023
Received in revised form
18 November 2023
Accepted 23 November 2023

Keywords:
Candidatus Accumulibacter
Comparative genomics
Horizontal gene transfer (HGT)
PhoU
Polyphosphate kinase 2 (PPK2)

ABSTRACT

Candidatus Accumulibacter, a prominent polyphosphate-accumulating organism (PAO) in wastewater treatment, plays a crucial role in enhanced biological phosphorus removal (EBPR). The genetic underpinnings of its polyphosphate accumulation capabilities, however, remain largely unknown. Here, we conducted a comprehensive genomic analysis of *Ca.* Accumulibacter-PAOs and their relatives within the Rhodocyclaceae family, identifying 124 core genes acquired via horizontal gene transfer (HGT) at its least common ancestor. Metatranscriptomic analysis of an enrichment culture of *Ca.* Accumulibacter revealed active transcription of 44 of these genes during an EBPR cycle, notably including the polyphosphate kinase 2 (PPK2) gene instead of the commonly recognized polyphosphate kinase 1 (PPK1) gene. Intriguingly, the phosphate regulon (Pho) genes showed minimal transcriptions, pointing to a distinctive fact of Pho dysregulation, where PhoU, the phosphate signaling complex protein, was not regulating the high-affinity phosphate transport (Pst) system, resulting in continuous phosphate uptake. To prevent phosphate toxicity, *Ca.* Accumulibacter utilized the laterally acquired PPK2 to condense phosphate into polyphosphate, resulting in the polyphosphate-accumulating feature. This study provides novel insights into the evolutionary emergence of the polyphosphate-accumulating trait in *Ca.* Accumulibacter, offering potential advancements in understanding the PAO phenotype in the EBPR process.

1. Introduction

With the rapid development of industry and the economy, there has been a significant surge in wastewater generation. This escalating wastewater production has, in turn, resulted in excessive phosphorus (P) discharge, leading to adverse consequences such as eutrophication, water quality deterioration, and aquatic ecosystem degeneration [1–3]. Enhanced biological phosphorus removal (EBPR) is an environmentally friendly and economical process widely applied in municipal wastewater treatment plants (WWTPs) for P removal [4–10]. This process is mediated by a group of microorganisms, namely polyphosphate-accumulating organisms (PAOs) [11–14]. *Candidatus* Accumulibacter is a model genus of PAOs commonly found in lab- and full-scale EBPR systems [15–18]. Under anaerobic conditions, *Ca.* Accumulibacter uses intracellularly stored polyphosphate (poly-P) as an energy source to power the uptake of volatile fatty acids (VFAs). This metabolic process results in the release of phosphate. The assimilated VFAs are then polymerized and stored as polyhydroxyalkanoates (PHAs). In the subsequent aerobic phase, PHAs are oxidized for cell metabolism and reproduction. Excess phosphate is removed from the aquatic phase to synthesize poly-P, achieving P removal [19–22]. This unique metabolic feature allows PAOs to thrive in alternating anaerobic-aerobic conditions, conferring sustainable P removal. However, the key genetic basis affording PAOs the ability to P

* Corresponding author. School of Environment and Energy, South China University of Technology, Guangzhou, 510006, China
** Corresponding author. Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, 637551, Singapore
E-mail addresses: SWuertz@ntu.edu.sg (S. Wuertz), qiugl@scut.edu.cn (G. Qiu).
[1] Authors contributed equally towards this study.

https://doi.org/10.1016/j.ese.2023.100353

cycling is unclear. Genes known to be indispensable for the P cycling feature, e.g., the polyphosphate kinase gene (*ppk*) and exopolyphosphatase gene (*ppx*) for poly-P synthesis and hydrolysis, respectively, and the inorganic phosphate transporter gene (*pit*) and the high-affinity phosphate transporter gene (*pst*) for phosphate transport, are widely preserved in the bacterial domain, including in non-PAOs [23,24]. Their presence does not guarantee the P cycling ability, and the key genes have yet to be identified. The transition from non-PAOs to PAOs may be driven by adaptive evolution [25,26].

The need to understand the gain and loss of genes in different strains and the genome diversification in a given lineage of organisms gave rise to pangenomics. A pangenome encompasses the entire set of genes from all individuals of a specific lineage [27,28]. Genes in a pangenome are divided into core genes and variable genes [29]. The collection of genes commonly present in all individuals of a specific lineage is called the core pangenome, representing the common genetic features of a microbial lineage [30]. The variable genes can be further divided into unique genes (found in a single strain/genome) and dispensable genes (shared in at least two but not all strains/genomes) [31]. Dispensable genes represent the intra-lineage diversity encoded among different members [29]. By avoiding single sample bias and ensuring full representation of genomic diversity of different lineage members, the analysis of the pangenome provides insight into the genetic basis of common phenotypic characteristics shared in a group of bacteria, greatly improving our ability to solve complex phenotypic problems [32–34]. Comparative genomics has been applied to study the evolution and development of many bacterial species [35–40]. Via comparative genomic analysis, Fernandez-Fueyo et al. [41] found a subset of potentially important genes for selective lignin decomposition in *Ceriporipsis subvermispora*.

Oyserman et al. [42] previously constructed a pangenome of the Rhodocyclaceae family (including ten *Ca.* Accumulibacter and 16 out-group genomes) to explore the genetic composition and evolutionary changes in metabolic pathways of the *Ca.* Accumulibacter genus. However, at the time, limited numbers of *Ca.* Accumulibacter genomes were available, with more than half having low completeness (<90%). The deficiency in genome quality and quantity may result in an inadequate representation of the lineage pangenome and affect the downstream analysis of genes. With the advance in EBPR research and sequencing techniques, increasing numbers of high-quality *Ca*. Accumulibacter genomes have been obtained [7,14,18,43–49]. New PAOs and glycogen-accumulating organisms (GAOs) were also identified in genera phylogenetically closely related to *Ca*. Accumulibacter. GAOs occupy a similar ecological niche as PAOs in EBPR systems. They use glycogen instead of polyphosphate as an energy source for anaerobic carbon source uptake, thus competing with PAOs. For instance, a *Propionivibrio* member was shown to perform as a GAO in full-scale WWTPs in Denmark [50]. Two *Dechloromonas* members in the same WWTPs (i.e., *Ca*. Dechloromonas phosphoritropha and *Ca*. Dechloromonas phosphorivorans) were revealed to be PAOs [51]. The identification of *Dechloromonas*-related PAOs raises the possibility that the emergence of the PAO phenotype may have occurred before the *Ca.* Accumulibacter's last common ancestor (LCA). The evolution in the P cycling feature needs to be re-evaluated and traced. Combined with the analysis of gene transcriptional characteristics of representative PAO strains, the key genomic characteristics distinguishing PAOs and non-PAOs may be further identified and determined, which would significantly advance the understanding of the genomic basis of the PAO phenotype.

To understand the emergence of the PAO phenotype of *Ca.* Accumulibacter, we selected 43 high-quality genomes within the Rhodocyclaceae family for comparative genomic analysis. A pangenome of the Rhodocyclaceae family, including 21 *Ca*. Accumulibacter genomes, seven of which were recovered from our EBPR reactors, 22 out-group genomes, including two confirmed *Dechloromonas* PAOs, i.e., *Ca*. Dechloromonas phosphoritropha and *Ca*. Dechloromonas phosphorivorans [51], and one *Propionivibrio* GAO genome, *Ca.* Propionivibrio aalborgensis [50], was constructed. In the analysis of genes within the pangenome, genes were classified as ancestral, derived, flexible, or lineage-specific genes. The dynamics in these genes in the evolutionary process were analyzed, and metatranscriptomic analyses were performed on an enrichment culture of *Ca*. Accumulibacter Clade IIC SCUT-2 for identifying their active genes in a typical anaerobic-aerobic cycle to narrow down the range of genes important for the PAO phenotype of *Ca*. Accumulibacter. Genomic comparisons were further performed between *Ca*. Accumulibacter, two *Dechloromonas*-related PAOs, and the *Propionivibrio* GAO. Among the numerous genes investigated, two key players emerged: the phosphate signaling complex protein gene (*pho*U) in the Pho regulon and the laterally derived polyphosphate kinase 2 gene (*ppk*2). These genes were identified as instrumental in the emergence of the PAO phenotype of *Ca.* Accumulibacter. This study provides new insights into the development of the P cycling trait of *Ca.* Accumulibacter.

2. Materials and methods

2.1. Data acquisition and evaluation

The genomes used for analysis included seven high-quality genomes recovered from our EBPR reactors and 36 genomes obtained from the National Center for Biotechnology Information (NCBI) database. All 43 genomes belong to the Rhodocyclaceae family, including 21 *Ca.* Accumulibacter genomes and 22 out-group genomes (ten *Dechloromonas*, seven *Thauera*, three *Azoarcus*, one *Propionivibrio*, and one *Zooglea ramigera* genomes). The completeness and contamination of the genomes were evaluated using CheckM [52]. The GenBank assembly accession, corresponding species names, and additional details about the qualities of these genomes can be found in the Supplementary Materials Tables S1–S3.

2.2. Orthologue analyses

Orthologous gene clustering is necessary for the reconstruction of the ancestral state. To find orthologous gene clusters based on the protein sequences, all vs. all BLAST of each Rhodocyclaceae genome was conducted using Orthofinder 2.5.4 [53] with parameters *-evalue 1e-5, -seg yes, -soft_masking true, -use_sw_tback*. The results were filtered to the query coverage ≥75% and the percent identity ≥70%. Orthologous gene clusters were identified using MCL version 14–137 with an inflation value of 1.1 [54].

2.3. Phylogenetic analysis of pangenome

Orthofinder was used to identify the pan single-copy genes for reliable phylogenic tree construction and gene flux analysis. The pan single-copy genes were aligned using the linsi option in MAFFT version 7.508 [55] and masked in Gblocks version 0.91b [56]. Seqkit (version 2.3.0) [57] was used to sort the single-copy gene sequences and convert the multi-line sequences into a one-line sequence. Iqtree version 2.2.0.3 [58] was used to predict the best phylogenetic tree model. Finally, the tree was constructed with model Q. insect+F+I+I+R4. Landscaping of the phylogenetic tree was achieved using iTOL version 6.6 [59].

2.4. Pangenome analysis

When a genome set has incomplete genomes, it is necessary to determine a threshold number of genomes in which a gene must be observed in order to call it 'core'. The probability that a gene was observed in all *Ca.* Accumulibacter genomes are the product of the completeness of each genome. The probability of a gene's absence in one genome while being present in all other genomes was computed by multiplying the completeness of the remaining genomes with the incompleteness (i.e., 1 minus the completeness) of the incomplete genomes. Cut-off values were calculated using the R script [60] (Supplementary Materials Spreadsheet 1). The maximum number of genomes allowing an effective calculation of the cutoff value was 21. Via a comprehensive evaluation of the quality and the clade distribution of all available genomes, 21 high-quality *Ca.* Accumulibacter genomes covering eight different clades were used for pangenomic analysis (The completeness and contamination of these genomes are documented in the Supplemental Materials Tables S1−S2).

2.5. Gene gain/loss analysis

Gene flux was analyzed using Count [61] based on the matrix of orthologous gene family abundance obtained in the previous analyses. For a more comprehensive examination of gene gain and loss dynamics, we applied a Wagner parsimony penalty of 2 [62,63]. Genes acquired before the node of the LCA of *Ca.* Accumulibacter were defined as ancestral, while those acquired at the node of *Ca.* Accumulibacter LCA were defined as derived genes. Genes determined to be obtained via horizontal gene transfer (HGT) in the derived genes were classified as laterally derived genes. Lineage-specific genes were present in a single *Ca.* Accumulibacter genome. Flexible genes were present in more than one but less than 18 *Ca.* Accumulibacter genomes. Genetic comparisons were performed between PAO and GAO genomes to understand the differences in their genetic makeup better. The pangenome is composed of 21 *Ca.* Accumulibacter and two *Dechloromonas* PAOs were denoted as the pan PAO genome. Core genes of the pan PAO genome were defined as genes belonging to the core genes of the pan *Ca.* Accumulibacter genome and were also present in two *Dechloromonas* PAO genomes. Differential genes were defined as core genes present in the pan PAO genome but absent in the *Ca.* Propionivibrio aalborgensis GAO genome.

2.6. Metabolic function analysis

The ancestral, derived, flexible, and lineage-specific genes were annotated and classified based on KEGG annotations [64] of clade IIC member SCUT-2 [49] and clade IIA member UW1 [4,26]. The number of genes annotated in each metabolic pathway was counted, with the number of each type of gene being divided by the total number of genes in the pathway. Metabolic pathways with high proportions of derived genes were considered to have undergone major changes during evolution.

2.7. Horizontal gene transfer (HGT) identification

Parametric and phylogenetic methods are commonly used to infer HGT [65]. This study used the phylogenetic method for HGT identification. Each derived gene was queried in the non-redundant (NR) database (published on May 7, 2015) [66] using the following BLASTP parameter [*-max_target_seqs 100-value 1E-6*] to preserve the first 100 BLAST results. The representative species were obtained from the first 100 BLAST results. Subsequently, the numbers and percentages of *Ca.* Accumulibacter, non-*Ca.* Accumulibacter Rhodocyclaceae, and non-Rhodocyclaceae members in the first 100 BLAST results were then calculated. A gene was considered a laterally derived gene if the numbers of *Ca.* Accumulibacter or non-*Ca.* Accumulibacter Rhodocyclaceae-related hits were less than 10%. All core and differential-derived genes in each metabolic pathway were analyzed to determine if they were obtained via HGT. The derived genes that were classified as HGT-originated are referred to as laterally derived genes. The origination of key genes (*ppk*2 and the homolog of *pho*U) was further confirmed using the phylogenetic method based on best-match analysis.

2.8. Metatranscriptomic analysis

An anaerobic-aerobic full-cycle study was performed on an enrichment culture of *Ca.* Accumulibacter Clade IIC SCUT-2 in the lab-scale EBPR reactor SCUT (Supplementary Materials). The P cycling activities and the transformation of carbon compounds were monitored. Activated sludge samples were collected just before the start of a sequencing batch reactor (SBR) cycle (0 min), and at 5 min (anaerobic phase), 30 min (anaerobic phase), 105 min (aerobic phase), and 120 min (aerobic phase) of the SBR cycle. The samples were snap-frozen in liquid N_2 and stored at −80 °C before the extraction of ribonucleic acid (RNA) for metatranscriptomic analysis.

For metatranscriptomic analysis, total RNA was extracted using the RNA PowerSoil® Total RNA Isolation Kit (Omega Bio-Tek, GA, USA). Fastp [67] and SortMeRNA [68] removed adaptation sequences and ribosomal ribonucleic acids (rRNAs). Filtered reads were mapped to the corresponding *Ca.* Accumulibacter draft genome (i.e., SCUT-2) using BBMap version 38.96 [69] and were normalized to transcript per million (TPM). Genes with TPM >100 were considered to be highly transcribed. Details on the reactor operation, full-cycle study, sample collection, metagenomic analysis, and metatranscriptomic analysis are found in the Supplementary Materials. Raw reads and draft genomes obtained were submitted to NCBI under BioProject No. PRJNA807832 and No. PRJNA771771.

3. Results

3.1. Identification of orthologous gene clusters

A total of 60722 pan Rhodocyclaceae orthologous gene clusters were identified, including 25080 homologous genes in the *Ca.* Accumulibacter pangenome (Supplementary Materials Spreadsheet 2, Sheets 1 and 3). Large proportions (63.8% and 54.7%) of gene families in the pan Rhodocyclaceae and pan *Ca.* Accumulibacter genomes were present in only single genomes (Fig. 2a and b). Approximately 1% (626) of gene families were present in ≥37 of the 43 genomes, which were used to define the core pan Rhodocyclaceae genome (Fig. 2c). In the pan *Ca.* Accumulibacter genome, 6.9% of genes were shared in ≥18 genomes (Fig. 2d). Non-paralogous genes (average gene copy per genome = 1) account for high proportions of pan Rhodocyclaceae and pan *Ca.* Accumulibacter genomes (95.6% and 93.8%, respectively) (Fig. 2e and f). The orthologous gene cluster identification results include the number of representative genes in each genome and summary statistics of pan Rhodocyclaceae and pan *Ca.* Accumulibacter gene clusters are provided in Supplementary Materials Spreadsheet 2 (Sheets 2 and 4).

3.2. Gene flux analysis

Among the 25080 gene clusters in the pan *Ca.* Accumulibacter genome, 2499 (9.96%) were inferred to occur in the genome of the

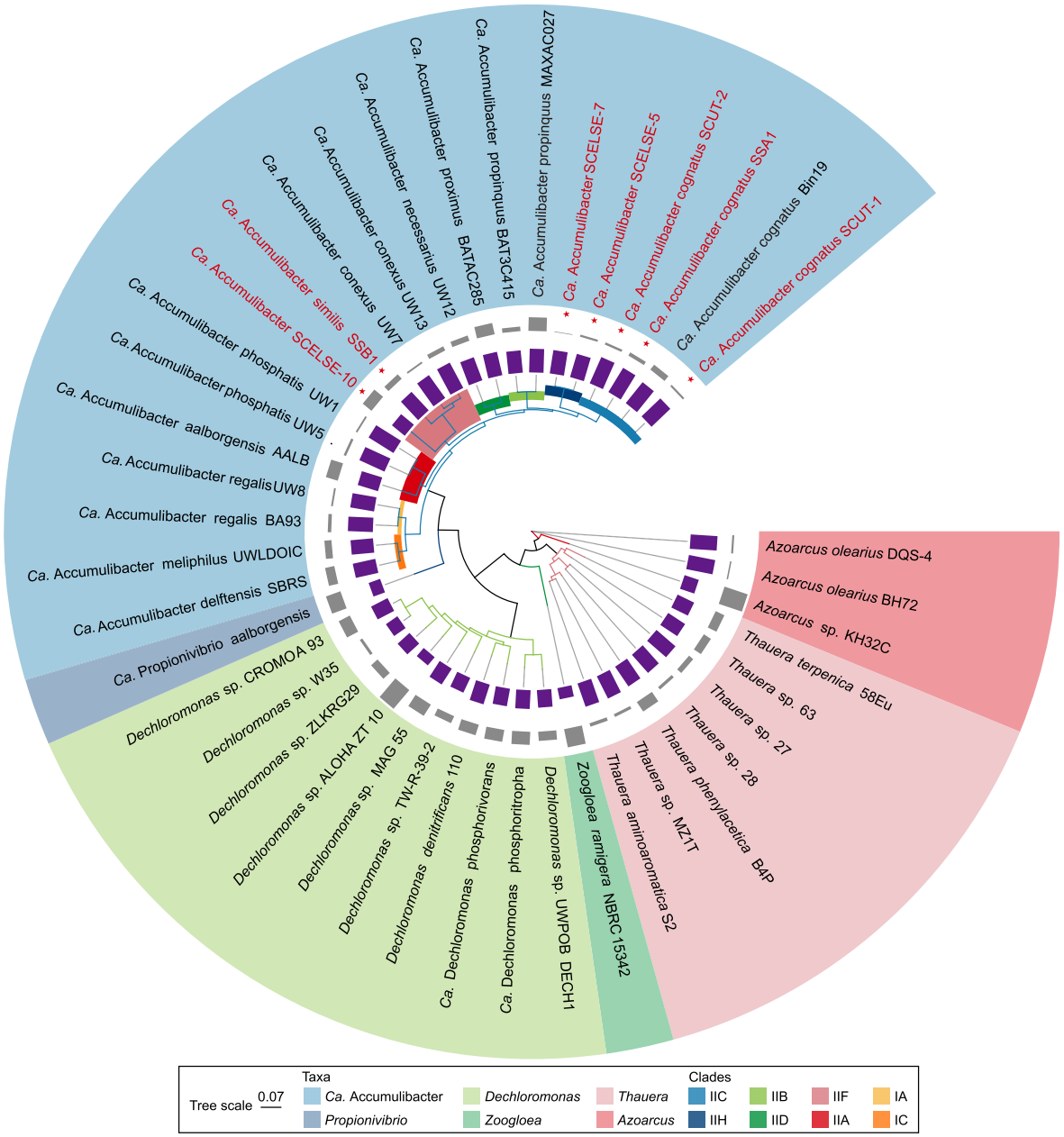

Fig. 1. A phylogenetic tree of 43 Rhodocyclaceae members was built based on the concatenation of 59 single-copy genes. The genomes in red were recovered from our lab-scale reactors. SSA1, SSB1, and SCUT-1 were recovered in our previous work [43,48]. SCELSE-5, SCELSE-7, SCELSE-10, and SCUT-2 were recovered from three of our lab-scale EBPR reactors (Supplementary Materials). The purple bars represent the number of shared orthogroups. The gray bars represent the number of unassigned genes.

LCA, and 1668 (6.73%) occurred before the LCA. Eight hundred eighteen (3.26%) were acquired at the node of LCA. Gene occurrence possibility calculation suggested that with a genome-number cutoff of 18, 99.94% of core genes could be identified (Fig. 3a). At this cutoff value, 1725 (6.88%) core genes were identified in the pan *Ca.* Accumulibacter genome (Fig. 3b and c). By further reducing the cutoff value to 17, the number of core genes increased from 1725 to 1829, and those with known functions increased from 298 to 318. As this study mainly focused on the changes in the genetic content, i.e., new core-derived genes and horizontally transferred genes, looser cutoff values did not seem to bring new gains. Thus, a relatively stricter cut-off value (i.e., 18) was used to ensure the accuracy of the results. The gene gain or loss of a pangenome needs to be characterized in specific lineage member genomes. To facilitate a subsequent combination with the transcriptome data, SCUT-2 and UW1 were used as representative genomes for gene flux analysis. Each gene in Clade IIC SCUT-2 and Clade IIA UW1 genomes was classified as ancestral, derived, lineage-specific, or flexible genes. There were no significant differences in the numbers and proportions of ancestral and flexible genes in these two genomes (ancestral genes accounted for 32.6% and 34.7%; flexible genes accounted for 43.8% and 43.6% in SCUT-2 and UW-1, respectively). Six hundred thirty eight and eight hundred and two derived genes were found in the SCUT-2 and UW1 genomes (17.6% and 14.0%, respectively). One hundred eighty nine lineage-specific genes (genes occurred only in UW1) were observed in UW1, which was slightly less than those (i.e., 275) in the SCUT-2 genome (Fig. 3d). Fig. 4 and Supplementary Materials Spreadsheet 3 provided

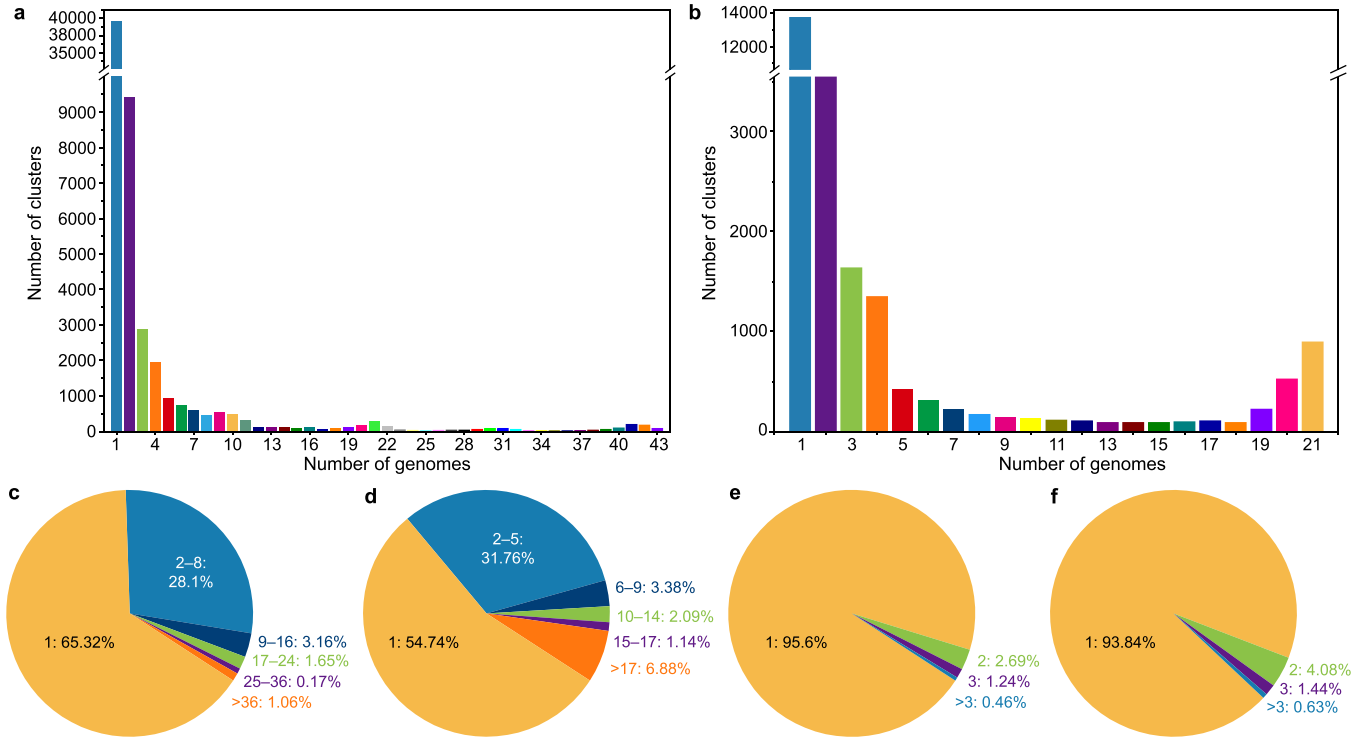

Fig. 2. **a**–**b**, The number of gene clusters at different frequencies in the pan Rhodocyclaceae genome (**a**) and the pan *Ca.* Accumulibacter genome (**b**). **c**–**d**, The proportion of clusters at different frequencies in the pan Rhodocyclaceae genome (**c**) and the pan *Ca.* Accumulibacter genome (**d**). **e**–**f**, The proportion of different average gene copies per genome in the pan Rhodocyclaceae genome (**e**) and the pan *Ca.* Accumulibacter genome (**f**). In each orthogroup, the average gene copies per genome are defined as the number of genes divided by the number of genomes.

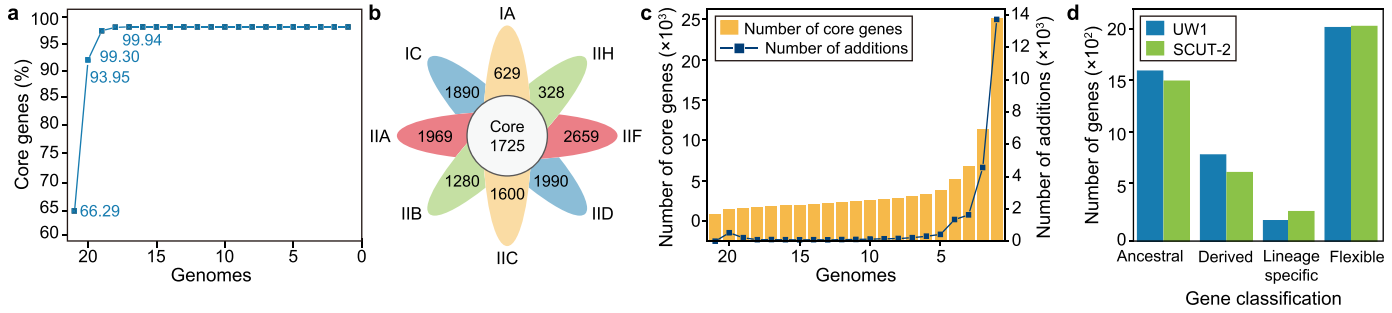

Fig. 3. **a**, Using the genome integrity estimate, about 99.94% of the core genes could be identified with a cut-off value of 18. Only gene families that appear in ≥18 *Ca.* Accumulibacter genomes are considered core genes. **b**, A Venn diagram describing the numbers of core genes and lineage-specific genes in the pan *Ca.* Accumulibacter genome. **c**, The number of core genes observed at different cutoff values. **d**, The number of genes assigned as ancestral, derived, lineage-specific, and flexible genes in SCUT-2 and UW1.

additional details about the presence, gain and loss of genes, and the discrete categories to which they were assigned.

3.3. Evolution of *Ca.* Accumulibacter metabolic pathways

The collections of genes identified as ancestral, derived, flexible, and lineage-specific genes were annotated using KEGG [64] and were grouped into different metabolic pathways. In SCUT-2, 2293 genes were annotated to various metabolic pathways. The translation metabolic pathway had the highest proportion of ancestral genes (77, accounting for 96%). The largest number of ancestral genes (224) and derived genes (63) was observed in the carbohydrate metabolism pathway, accounting for 63% and 18%, respectively. The highest proportion (15 out of 53, 28.0%) of derived genes was observed in the cell growth and death metabolic pathway (Fig. 5a). Within each primary pathway, ancestral and derived genes also showed distinct proportions in different secondary pathways. For instance, within the carbohydrate metabolism, the galactose metabolism pathways had the highest proportion (4 out of 5, 80%) of derived genes. Whereas ancestral genes dominated the citric acid cycle (TCA cycle) (25 out of 30, 83%) and the glyoxylate and dicarboxylate metabolism pathways (33 out of 45, 73%). In signal transduction, the two-component system contained the highest proportion of derived genes (27 out of 182, 15%). In membrane transport, among the 122 ABC-transporter encoding genes, 18 were derived (15%) (Fig. 5). Similar number and proportion of genes assigning to various metabolic pathways were observed in the *Ca.* Accumulibacter clade IIA UW1 genome with only two metabolic pathways (transport and catabolism, cell growth and death) showing significant differences in the proportions of derived genes (28% and 40% in SCUT-2 and 14% and 23% in UW1, respectively) (Fig. 5 and Supplementary Materials Fig. S1). These results

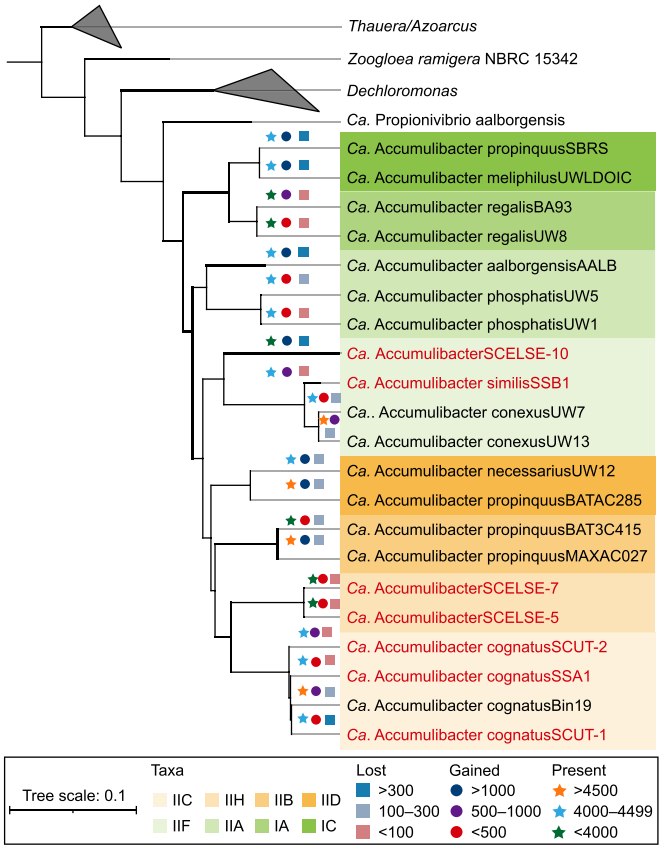

Fig. 4. Gain or loss of genes at various nodes of the *Ca.* Accumulibacter lineage. A maximum likelihood tree was built based on the concatenation of single-copy genes with model Q. insect+F+I+I+R4. Genomes in red are those recovered from our bioreactors [7,43,48,49].

indicated that different strains of *Ca.* Accumulibacter underwent comparable developmental changes during evolution but, at the same time, preserved a certain degree of gene diversity. Detailed annotation of each gene in SCUT-2 and UW1 can be viewed in Supplementary Spreadsheet 4.

3.4. Pan *Ca.* Accumulibacter phylogenetic analysis of derived genes

Relatively strict parameters (i.e., 70% identity and 75% coverage) were used to identify homologous gene clusters. The derived genes were manually classified into those derived from accumulative mutations and HGT. Phylogenetic analysis was further performed to confirm that *ppk*2 and the homolog of *pho*U are horizontally derived (Supplementary Materials Fig. S4). Among 298 core-derived genes successfully annotated in KEGG, 124 were shown to have been acquired via HGT. Among the 124 genes, 67 were involved in KEGG pathways. The carbohydrate metabolism pathway harbors the highest numbers (25) of derived genes via HGT, including these in glycolysis/gluconeogenesis (e.g., genes encoding the phosphoglucomutase, the glucokinase, and the phosphoglycerate kinase), starch and sucrose metabolism (e.g., the starch synthase, and the glycogen phosphorylase genes), and in butanoate metabolism (genes encoding the poly[(R)-3-hydroxyalkanoate polymerase subunits). In signal transduction, the two-component system contained ten laterally derived genes, such as genes encoding the REDOX signal transduction system proteins RegA/B and the phosphate regulon proteins PhoR-PhoB. Another remarkable set of genes derived via HGT was oxidative phosphorylation in the energy metabolism pathway, including these encoding the NADH-quinone oxidoreductase subunit, the polyphosphate kinase, and the cytochrome C. The inorganic phosphate transporter gene (*pit*) was also acquired via HGT. Similar results were observed for UW1. In the two-component system, genes encoding the REDOX signal transduction system proteins RegA/B and the phosphate regulon proteins PhoR-PhoB were laterally derived. More details about the BLAST comparison results can be found in Supplementary Materials Spreadsheet 5.

3.5. Comparison of genetic compositions in PAOs and non-PAOs

In the context of our investigation, the presence or absence of specific genes in *Ca.* Accumulibacter, compared to closely related PAOs and non-PAOs, holds significant implications for elucidating the genetic basis of the P cycling phenotype. If a gene was present in *Ca.* Accumulibacter, but absent in other closely related PAOs, may also not be a key to developing the P cycling phenotype. Conversely, if a gene was present in *Ca.* Accumulibacter, or their closely related PAOs but absent in non-PAOs, might be a key gene to the emergence of the PAO phenotype. For a better understanding of the genomic difference between closely related PAOs and non-PAOs, a pan PAO genome (composed of 21 *Ca.* Accumulibacter and two *Dechloromonas* PAOs) [51] analysis was performed. The pan PAO genome was compared to the *Ca.* Propionivibrio aalborgensis (a closely related GAO) [50] genome to identify differential genes (defined as core genes present in the pan PAO genome but absent in the *Ca.* Propionivibrio aalborgensis genome). In the pan PAO genome, 124 differential genes were identified. Alkaline phosphatase synthesis response regulator (PhoP) and polyphosphate kinase 2 (PPK2) genes were both differential genes. Other genes in the operon or the genes regulated by PhoP were not differential genes. Carbohydrate metabolism had the largest differential genes (16), including those encoding the acetyl-CoA C-acetyltransferase and the enoyl-CoA hydratase. The cofactor and vitamin metabolic pathway harbored the second largest number of differential genes (11), followed by energy metabolism (9), replication and repair (6), and signal transduction (5) metabolic pathways. The lowest number (1) of differential genes was observed in the transcription and metabolism of other amino acid pathways. Further analysis of another 21 available *Propionivibrio* genomes confirmed that *ppk*2 and *pho*U are differential genes between *Ca.* Accumulibacter and *Propionivibrio*. HGT analysis was aimed at gene acquisition in *Ca.* Accumulibacter during evolution, based on the hypothesis that the emergence of the P cycling ability by PAOs resulted from the acquisition of certain key genes. However, the hypothesis ignored the possibility that non-PAOs may have lost certain key genes in the process of evolution, leading to their inability to remove P. Differential genes included gene loss in non-PAOs during evolution. The analysis in this part allows us to understand the evolutionary process from a different perspective more comprehensively. More details about the differential genes (metabolic pathway and functional annotation) can be found in Supplementary Materials Spreadsheet 6.

3.6. Metatranscriptomic profiles

By analysis of the gene transcription levels of *Ca.* Accumulibacter in a typical EBPR cycle, we excluded genes that displayed no remarkable transcription in the comparative genome may be excluded. Thus, the range of genes could be further narrowed down, facilitating the identification of key genes important to the PAO phenotype. Metatranscriptomic analysis was performed on an enrichment culture of *Ca.* Accumulibacter clade IIC strain SCUT-2 (with a relative abundance of 37.1%, as suggested by the

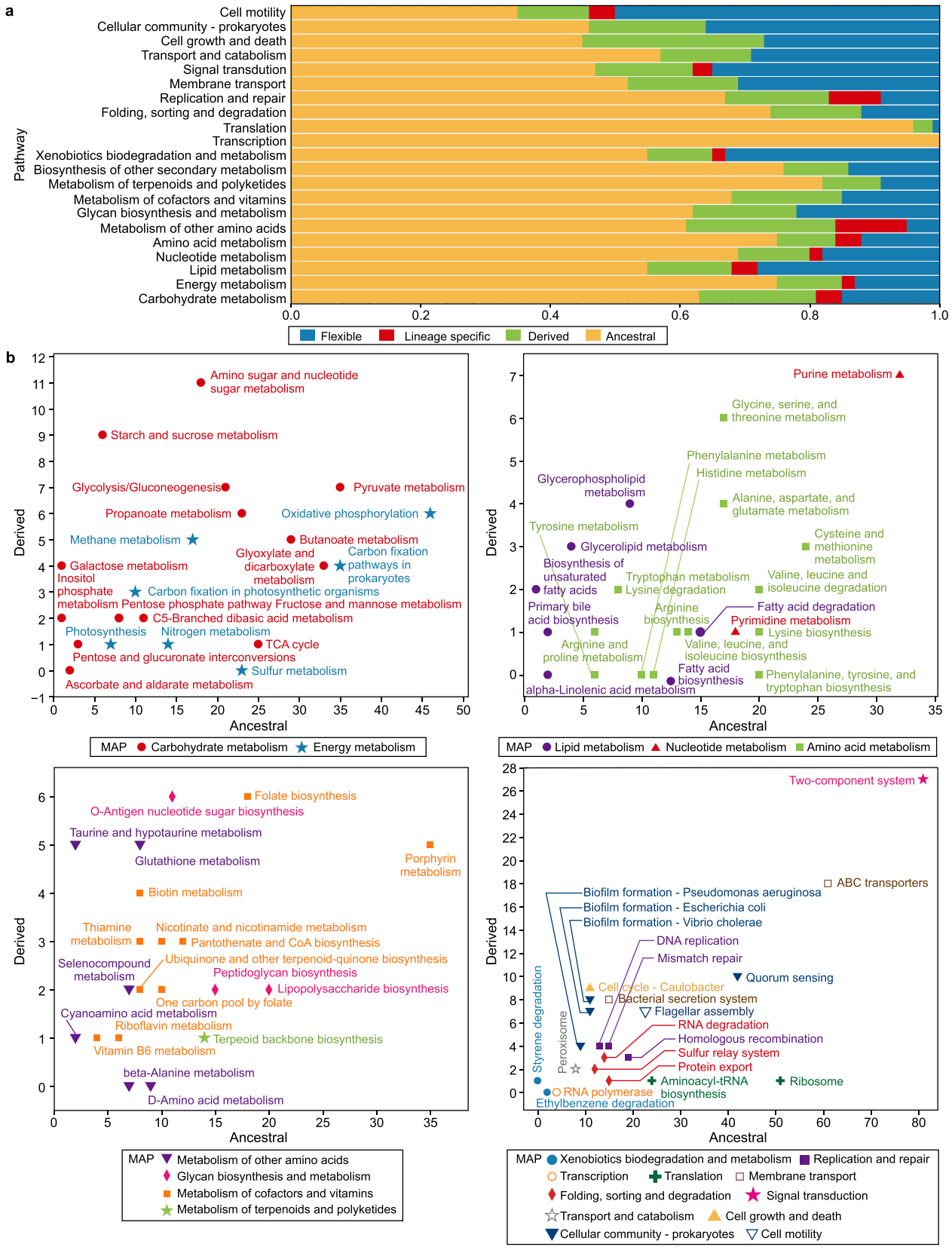

Fig. 5. **a**, The ratio of ancestral, derived, lineage-specific, and flexible genes in different primary metabolic pathways (MAP) of SCUT-2. **b**, The number of ancestral and derived genes in representative secondary MAP of SCUT-2.

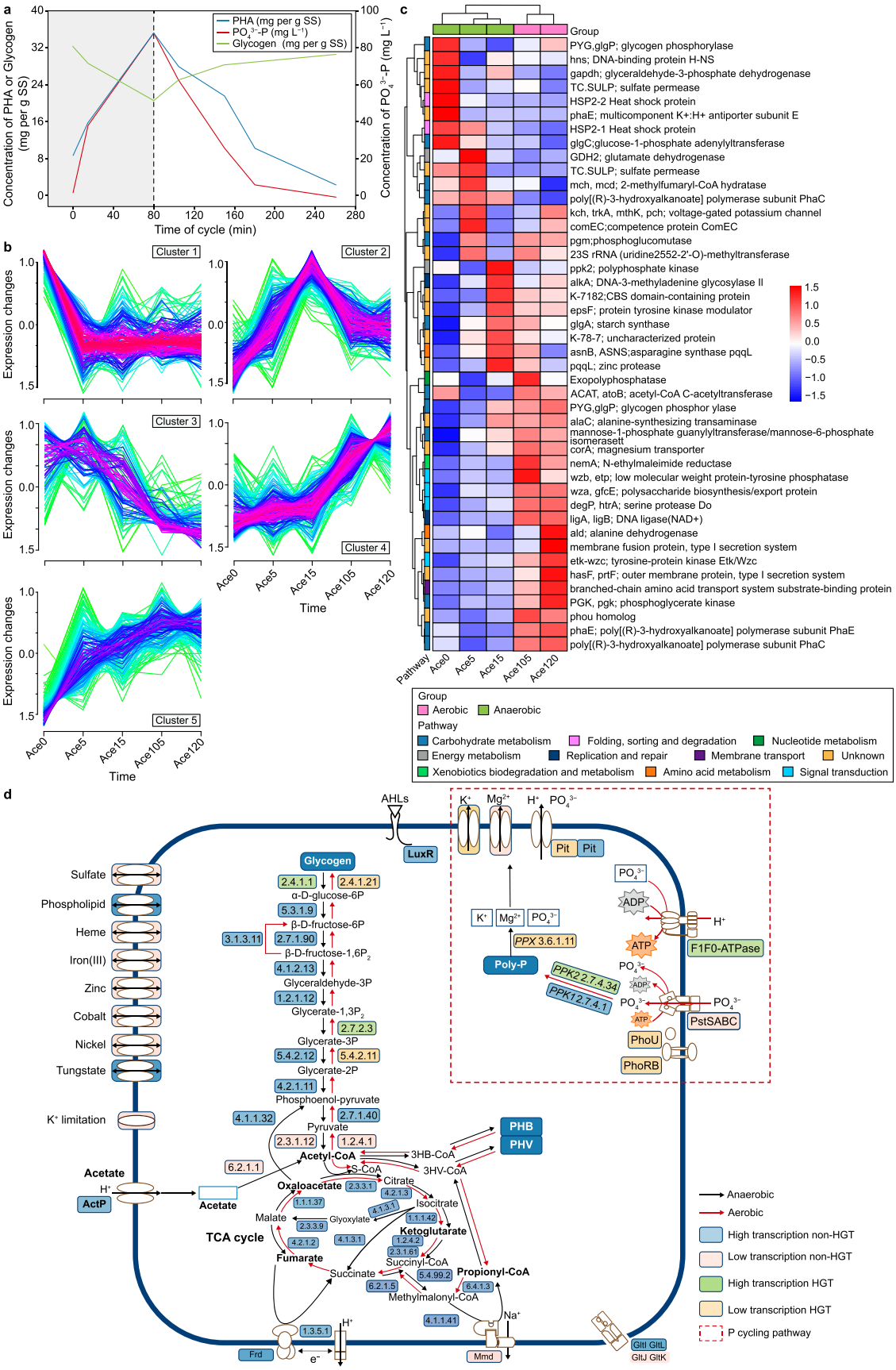

Fig. 6. **a**, Changes in phosphate, PHA, and glycogen concentrations during an anaerobic-aerobic full cycle. **b**, Cluster analysis of transcriptome data at different time points for transcription pattern identification. **c**, 44 highly transcribed and laterally derived genes (via HGT) in the SCUT-2 genome during the anaerobic-aerobic full cycle.

(continued on next page)

metagenomic analysis). In the SCUT-2 genome, out of 5037 annotated genes, 906 were highly transcribed (TPM >100). There were 298 core-derived genes, 84 of which were highly transcribed (Supplementary Materials Spreadsheet 7). To understand the dynamic patterns and functional relationships of 905 core genes with known function, they were classified into five clusters using the Mfuzz [70] (Fig. 6b). Most genes (e.g., the acetate permease gene *act*P, NOF05_02545) in Cluster 1 were related to the transporter for carbon uptake and energy utilization. Cluster 2 showed a pattern of increased transcription throughout the anaerobic period, peaking after oxygen exposure. Key members of this cluster included the phosphate transport system substrate-binding protein (*pstS*, NOF05_04305) and the laterally derived polyphosphate kinase 2 gene (*ppk2*, NOF05_17285). Cluster 3 genes showed high transcription at the beginning of the anaerobic stage and reduced towards the end of the anaerobic cycle, correlating with the depletion of acetate (Fig. 6a). Their high transcription in the aerobic stage was mostly related to the routing of anaerobically stored carbon to the TCA cycle and glycogenesis [7,26]. Cluster 4 contained genes encoding the distant homolog of PhoU (NOF05_17860, NOF05_12350) and antitoxin CptB (NOF05_13125), which showed low transcription during the anaerobic stage but were upregulated during the aerobic phase. These genes possibly play a role in sustaining vital activities and controlling homeostatic environments [71]. Finally, genes in Cluster 5 may be associated with the maintenance of stable intracellular environments or cell growth, including genes encoding the ion transporters, such as the magnesium transporter gene (NOF05_18175) and the low-affinity inorganic phosphate transporter (*pit*, NOF05_12345). These clustering patterns aligned with the metabolic characteristics of *Ca.* Accumulibacter in EBPR (Fig. 6a).

The transcription of horizontally transferred genes in SCUT-2 was further analyzed. 44 genes, which were identified to be obtained via HGT, were highly transcribed (Fig. 6c). These genes were involved in pathways, such as glycolysis/gluconeogenesis (phosphoglycerate kinase, and phosphoglucomutase), ABC transporters (branched-chain amino acid transport system substrate-binding protein), butanoate metabolism (poly[(R)-3-hydroxyalkanoate] polymerase subunit), the two-component system (low molecular weight protein-tyrosine phosphatase, polysaccharide biosynthesis/export protein, tyrosine-protein kinase, and serine protease), transporters for inorganic salts (sulfate permease, and magnesium transporter), and showed high transcription throughout the EBPR cycle. Polyphosphate kinase 2 gene (*ppk*2) was also highly transcribed and was significantly upregulated in the anaerobic phase. The transcription of the phosphate transport regulator (a distant homolog of PhoU) was significantly upregulated in the aerobic stage. PHA synthesis-related genes were also highly transcribed. A full list of the SCUT-2 gene transcription data can be found in Supplementary Materials Spreadsheet 7.

Comparisons were further made to the gene transcription characteristics of UW1 [26]. 35 horizontally derived gene families were highly transcribed in both SCUT-2 and UW1 (Supplementary Materials Fig. S2). Apart from the homolog *pho*U genes and *pit*, which are related to phosphate regulation and transport, 42 laterally derived gene families were under-transcribed in SCUT-2 but highly transcribed in UW1, including the acetate kinase gene. These 42 gene families may not play a key role in the evolution of non-PAO to PAO due to their different transcription behaviors in SCUT-2 and UW1. Combined with transcriptomic analysis, the range of key genes can be effectively reduced, and a metabolic model of *Ca.* Accumulibacter can be constructed (Fig. 6d). Most genes in the central carbon metabolic pathway were highly transcribed non-HGT genes, indicating that this pathway is indispensable for *Ca.* Accumulibacter, yet raises doubts about its direct involvement in the evolution from a non-PAO metabolism to a PAO. In the P cycling pathway, several laterally acquired genes were involved, suggesting their potential pivotal role in the evolution of *Ca.* Accumulibacter. Some of them were highly transcribed, further implying their importance in the evolution of *Ca.* Accumulibacter (Fig. 6).

4. Discussion

Previous research suggested that the transition of PAO from non-PAO may have occurred at the node of *Ca.* Accumulibacter LCA [42]. However, a recent investigation has put forth compelling evidence indicating the presence of PAOs in the *Dechloromonas* genus (i.e., *Ca.* Dechloromonas phosphoritropha, *Ca.* Dechloromonas phosphorivorans) [51], raising a possibility that the emergence of the PAO phenotype may have occurred before the *Ca.* Accumulibacter LCA. Here, we discuss the function of key laterally derived genes in the context of pangenomics and known PAO metabolism. A metatranscriptomic analysis of an enrichment culture of *Ca.* Accumulibacter Clade IIC member SCUT-2 contrasting those of *Ca.* Accumulibacter Clade IIA UW1 was performed to study the transcriptional dynamics of key genes in *Ca.* Accumulibacter. This approach allowed the exclusion of genes that were not highly transcribed in the large collection of laterally derived genes to narrow down the range of key genes to obtain new insights on key genomic features of the polyphosphate accumulating trait.

4.1. Carbon substrate uptake

The largest number of genes were annotated to the carbohydrate metabolism pathway in both SCUT-2 and UW1 genomes (354 and 369, respectively). The SCUT-2 genome contained 224 ancestral genes, 63 derived genes, and 49 laterally derived genes. Transcriptomic analysis suggested that when acetate was used as a carbon source, genes directly related to intracellular acetate processing and PHA synthesis were remarkably upregulated in SCUT-2 (Supplementary Materials Spreadsheet 7). The high-affinity acetyl-CoA synthetase (NOF05_02565) and low-affinity phosphate acetyltransferase (NOF05_11790) are responsible for acetate activation [11,72]. Other genes involved in the acetyl-CoA pathway, including the pyruvate kinase gene (NOF05_14290) and the phosphoenolpyruvate carboxykinase gene (NOF05_14615), maintained high levels of transcription throughout the anaerobic-aerobic cycle. However, these genes are all ancestral genes. Only one horizontally transferred gene (i.e., the acetate kinase gene, NOF05_16845) was barely transcribed. Therefore, genes related to acetate processing may not be pivotal factors contributing to the emergence of the PAO phenotype. In addition, in the TCA cycle [73], there were 30 genes. Among them, only the dihydrolipoamide dehydrogenase gene (NOF05_18520) was laterally derived, whereas transcribed at a low level. This indicates that the gain/loss of genes in the TCA cycle might not have contributed remarkably to the evolution of non-PAOs to PAOs. Four laterally derived genes occurred in the PHA synthesis pathway (*pha*C NOF05_18015, NOF05_21650,

Fig. 6 (continued)
d, A metabolic model of *Ca*. Accumulibacter. Black and red solid arrows represent active metabolic pathways in the anaerobic and aerobic phases. Genes in blue and pink are genes not acquired via HGT with high and low transcription, respectively. Genes in green and yellow represent genes acquired via HGT with high and low transcription, respectively. The red dashed line denotes the key P cycling pathway. The enzyme commission (EC) number indicates the key enzyme involved in each pathway/reaction.

NOF05_21620, and *pha*A NOF05_18020), NOF05_21650 and NOF05_21620 were highly transcribed throughout the EBPR cycle (Fig. 6). Whereas *Ca.* Propionivibrio aalborgensis also encoded these genes [50]. Their contribution to the evolution from a non-PAO metabolism to a PAO metabolism was unlikely.

4.2. Two-component systems

The two-component signal transduction system enables bacteria to sense, respond, and adapt to diverse and dynamic environmental conditions [74]. This system is commonly preserved in the bacterial domain. The number of genes in the two-component system was considered to be closely related to the bacteria's living environment [75]. Bacteria living in extreme environments tend to encode many signaling proteins for improved adaption [76]. In the SCUT-2 genome, a total of 182 genes were annotated to the two-component system, including 81 ancestral genes and 27 derived genes. Notably, 12 of these genes have been acquired laterally. In both SCUT-2 and UW1, phosphate regulon response regulator gene *pho*B (NOF05_18105), phosphate regulon sensor histidine kinase gene *pho*R (NOF05_18105), and redox signaling genes *reg*A and *reg*B (NOF05_11115, NOF05_11120) were laterally derived. RegB/RegA was shown to control and regulate a variety of basic metabolic processes in *Rhodobacter*, *Capsulatus*, and *Sphaeroides*, such as photosynthesis, CO_2 fixation, N_2 assimilation, denitrification, and electron transport [77] via direct or indirect control of respective operons [78,79]. However, both *reg*A and *reg*B were absent in two *Dechloromonas* PAO genomes (GCA_016722705.1 and GCA_016721185.1) [51], suggesting that the redox signaling RegA/B were not indispensable for a PAO phenotype. PhoR-PhoB is present in both *Ca.* Accumulibacter and two *Dechloromonas* PAO genomes can potentially play a role in PAO phenotype evolution. Since the PhoR-PhoB system is a part of the Pho regulon, further discussion was provided in the following subsection.

4.3. Phosphate regulatory system

The phosphate regulator (Pho) is a regulatory mechanism to maintain and manage inorganic phosphate concentrations in bacterial cells. The system typically consists of extracellular enzymes, transporters, and enzymes involved in the intracellular storage of phosphate [80]. Signal transduction of Pho regulators requires seven proteins, including PhoR, PhoB, four components of the ABC transporter Pst (PstS, PstA, PstB, and PstC), and PhoU. An increase in the extracellular phosphate concentration near the PstSCAB transporter would increase phosphate binding to PhoU, inhibiting the PhoR kinase activity and the PstSCAB transporter activity. In the absence of phosphate input, PhoU dissociates with phosphate, allowing the phosphate transport (Pst) to return to a normal working state [81]. The above feedback control enables bacteria to maintain and control a relatively stable intracellular phosphate concentration. Most of the genes in the Pho regulatory system in *Ca.* Accumulibacter are laterally derived, including those encoding PhoR and PhoB. In addition, within the genomes of *Ca.* Accumulibacter, three distant homologs of the phoU gene (designated as NOF05_17860, NOF05_09930, and NOF05_09935) were identified. Distant homologs are protein pairs with similar structures and functions but low gene sequence similarity [82]. The homolog *pho*U is located in the *pit* operon within *Ca.* Accumulibacter genomes. Moreover, PhoR-PhoB is also present in two *Dechloromonas* PAO genomes (*Ca.* Dechloromonas phosphoritropha and *Ca.* Dechloromonas phosphorivorans).

In SCUT-2, the transcription of *pho*R (NOF05_18110) and *pho*B (NOF05_18105, NOF05_19100) was negligible. The transcription level of *pho*R (CAP2UW1_1997) in UW1 was also low. The transcription of *pho*B (CAP2UW1_1996) in the aerobic phase was slightly upregulated (with TPM values from 12 to 92) but was still at relatively low levels (Supplementary Materials Spreadsheet 5). These results suggest that PhoR-PhoB in *Ca.* Accumulibacter was probably not active in perceiving phosphate concentrations.

Similarly, the *pho*U genes were almost not transcribed (with the maximum TPM values < 12, Fig. 6). Although the homolog *pho*U genes showed high transcription, the trend was not in line with *pst*, indicating that the PhoU or their laterally derived homologs were not effectively regulating Pst (Supplementary Materials Fig. S3). The same phenomenon was observed in UW1 [26] and UW6 [45] metatranscriptome (Supplementary Materials Spreadsheet 7). In *Staphylococcus aureus*, the absence of *pho*U homolog, located in the *pit* operon, leads to the upregulation of phosphate transporter genes (*pst*), increasing intracellular polyphosphate levels [71]. In *Sinorhizobium meliloti*, the absence of *pho*U resulted in excessive accumulation of phosphate, which inactivates cells due to P poisoning, resulting in poor cell growth [83,84]. Based on these results, we proposed two hypotheses. (1) PhoU in *Ca.* Accumulibacter was ineffective in regulating Pst even under high intracellular phosphate concentrations (no transcription of the *pho*U, and the unmatched transcription of *pho*U homolog and *pst*, Supplementary Materials Fig. S3). Pst continued to operate (as indicated by the high transcription of *pst* in the transcriptome, Supplementary Materials Fig. S3), resulting in excessive phosphate accumulation in cells (Fig. 6a). The laterally derived PPK2 functioned (as suggested by the high transcription of *ppk*2, Supplementary Materials Fig. S3) to condense excess phosphate into poly-P to avoid P poisoning. The second is that, in *Ca.* Accumulibacter, since *pho*U, the homolog of *pho*U and *ppk*2 were derived from different donor bacteria (Rhodocyclaceae, *Burkholderia*, and Gramaproteobacteria, respectively, as suggested by the BLAST results, Supplementary Materials SpreadSheet 5), their encoding proteins (i.e., PhoU, PhoU homolog, and PPK2) may have incompatible phosphate activation/inactivation thresholds. PPK2 continued to synthesize poly-P by consuming intracellular phosphate transported via Pst, resulting in consistently low intracellular phosphate concentration, which was insufficient to combine with PhoU and/or its homologs to downregulate Pst. In the SCUT-2 and UW1 transcriptomes (Fig. 6), PPK2 showed high levels of transcription during the entire EBPR cycle (with TPM values up to 12481 in SCUT-2), which was further up-regulated in the aerobic stage, suggesting that PPK2 worked to synthesize poly-P by consuming phosphate which was imported via Pst, avoiding possible cell inactivation and poisoning due to elevated intracellular phosphate concentrations and achieved poly-P accumulation. In addition, *Ca.* Dechloromonas phosporitropha lacked *pst*, *pho*U, *pho*B, and *pho*R genes in the Pho regulon, which is consistent with our hypothesis that the Pho regulation may not work properly in PAOs. The transcriptomics data of *Microlunatus phosphovorus* (BioProject No. PRJNA984968) and proteomics data of *Tetrasphaera elongate* (obtained from Herbst et al. [85]) were further analyzed to check whether the same mechamism could apply to other PAOs. In the transcriptome of *Microlunatus phosphovorus*, we found that the transcriptional patterns of *pst* were also inconsistent with those of *pho*U during an anaerobic and aerobic cycle (Supplementary Materials). From the proteome of *Tetrasphaera elongate*, the relative abundances of Pst and PhoU did not vary significantly between anaerobic and oxic conditions; hence, they were not significantly affected by changes in phosphate concentrations [85]. Taken together, these results suggest that in *Microlunatus phosphovorus* and *Tetrasphaera elongate*, the regulation of Pst by PhoU was not effective and that the Pho dysregulation mechamism may also apply to non-*Ca.* Accumulibacter PAOs. However, additional work is needed to confirm its broad applicability.

Despite that, there is limited research on the Pho regulatory system in *Ca.* Accumulibacter, the transcriptomics and gene origination analysis in the Pho regulon suggested that it may represent a key link in the emergence of the PAO phenotype.

4.4. Transport of phosphate

Phosphorus (organic and/or inorganic) is a typical restricting nutrient. Therefore, microorganisms developed adaptive mechanisms to cope with ordinary P deficiency. Low-affinity inorganic phosphate transport systems (Pit) and high-affinity phosphate transport systems (Pst) are key transporters used for inorganic phosphate transport [86,87]. In the pan *Ca.* Accumulibacter genomes, genes encoding the Pst transporter, are neither core nor laterally derived. Furthermore, *Ca.* Dechloromonas phosporitropha (PAO) do not encode any *pst* [51]. These results suggested that the Pst transport system may not be indispensable for a PAO phenotype. *Ca.* Dechloromonas phosporitropha encoded a phosphonates/phosphate transport system (Phn), which was shown to be a high-affinity phosphate transporter in *Mycobacterium smegmatis* [88]. This system may serve as a backup for the Pst transport system in *Ca.* Dechloromonas phosporitropha.

In the pan PAO genome, the low-affinity inorganic phosphate transporter gene (*pit*, NOF05_09925, NOF05_09940) was laterally derived. The efflux of phosphate in symport with H^+ via Pit produces proton motive force, which is a key driving force for the uptake of VFAs, lactate, succinate and amino acids by *Ca.* Accumulibacter [7,8,89]. Therefore, *pit* is an important feature gene for the PAO phenotype. In SCUT-2 transcriptomes, the transcription of the *pit* was upregulated during the transition from anaerobic to aerobic conditions (Supplementary Materials Fig. S3). The confirmed GAO, *Ca.* Propionivibrio aalborgensis, which is closely related to *Ca.* Accumulibacter (Fig. 1), are lack of *pit*. But *pit* is present in the genomes of other GAOs, for example, *Defluviicoccus* GAO-HK [90], *Ca*. Competibacter denitrificans, and *Ca.* Contendobacter odensis [91]. In addition, we analyzed 21 *Propionivibrio* genomes in the NCBI database. Pit transporter was encoded in 13 of 21 *Propionivibrio* genomes (Supplementary Materials Table S4). Anyhow, *pit* may not be a key feature driving the evolution of non-PAO into PAOs and may neither be used as a marker gene for the PAO phenotype, although it is indispensable for the P cycling trait.

5. Conclusion

In this study, we conducted pangomics with metatranscriptomic analysis on an enrichment culture of *Ca.* Accumulibacter clade IIC member SCUT-2. The primary objectives of this investigation were to understand the genomic transition in the evolution of *Ca.* Accumulibacter and to identify the key genes responsible for the emergence of the P-accumulating traits. Our study has brought forth several noteworthy findings.

(1) A total of 298 core genes were identified as novel acquisitions in the ancestral lineage of *Ca.* Accumulibacter, with 124 of them being derived via HGT. Notably, 44 of these laterally derived core genes were highly transcribed in a typical EBPR cycle.
(2) A high-affinity phosphate transport system (Pst) may not be indispensable for the PAO phenotype. Inorganic phosphate transporter (Pit) may not be a key feature driving non-PAO evolution into PAOs. Consequently, their encoding genes may not be reliable markers for the PAO phenotype.
(3) Low transcription of the *pho*R-*pho*B two-component system genes and the unmatched transcription of *pst* and *pho*U implied that the Pho regulon may not function properly in *Ca.* Accumulibacter.
(4) A Pho dysregulation mechamism was proposed. The PhoU and laterally derived PhoU homologs in *Ca.* Accumulibacter were ineffective in regulating Pst, resulting in excessive P uptake. To avoid P poisoning, the laterally derived PPK2 was employed to condense excess phosphate into poly-P. Alternatively, PhoU and PPK2 genes were derived from different donor bacteria, resulting in unmatched activation/inactivation thresholds. PPK2 tends to reduce the intracellular phosphate concentration to levels perceived by PhoU as low-phosphate states, thereby promoting continuous phosphate uptake.

This study is expected to provide a new perspective for understanding the development and evolution of the P cycling traits for *Ca.* Accumulibacter.

CRediT authorship contribution statement

Xiaojing Xie: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization. **Xuhan Deng:** Data Curation, Resources, Visualization. **Liping Chen:** Data Curation, Resources, Visualization. **Jing Yuan:** Investigation, Resources, Data Curation. **Hang Chen:** Investigation, Resources, Data Curation. **Chaohai Wei:** Writing – Review & Editing, Supervision. **Xianghui Liu:** Investigation, Resources, Data Curation. **Stefan Wuertz:** Supervision, Writing – Review & Editing, Project Administration, Funding Acquisition. **Guanglei Qiu:** Conceptualization, Methodology, Investigation, Supervision, Writing – Review & Editing, Validation, Project Administration, Funding Acquisition.

Data available

All data generated or analyzed during this study are included in this published article. Metagenomic raw reads and draft genomes were submitted to NCBI under BioProject No. PRJNA807832 and No. PRJNA771771. Metatranscriptomic data were submitted to NCBI under the submitted No. PRJNA807832. Other data were documented in the Supplementary Materials.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (52270035 and 51808297), the Natural Science Foundation of Guangdong Province (2021A1515010494), the Guangzhou Key Research and Development Program (2023B03J1334), and the Pearl River Talent Recruitment Program (2019QN01L125).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2023.100353.

References

[1] J.T. Bunce, E. Ndam, I.D. Ofiteru, A. Moore, D.W. Graham, A review of

phosphorus removal technologies and their applicability to small-scale domestic wastewater treatment systems, Front. Environ. Sci. 6 (2018) 8.
[2] A. Abdelfattah, S.S. Ali, H. Ramadan, E.I. El-Aswar, R. Eltawab, S.-H. Ho, T. Elsamahy, S. Li, M.M. El-Sheekh, M. Schagerl, M. Kornaros, J. Sun, Microalgae-based wastewater treatment: mechanisms, challenges, recent advances, and future prospects, Environmental Science and Ecotechnology 13 (2023) 100205.
[3] G. Qiu, Y. Law, R. Zuniga-Montanez, X. Deng, Y. Lu, S. Roy, S.S. Thi, H.Y. Hoon, T.Q.N. Nguyen, K. Eganathan, X. Liu, P.H. Nielsen, R.B.H. Williams, S. Wuertz, Global warming readiness: feasibility of enhanced biological phosphorus removal at 35 °C, Water Res. 216 (2022) 118301.
[4] H.G. Martín, N. Ivanova, V. Kunin, F. Warnecke, K.W. Barry, A.C. McHardy, C. Yeates, S. He, A.A. Salamov, E. Szeto, E. Dalin, N.H. Putnam, H.J. Shapiro, J.L. Pangilinan, I. Rigoutsos, N.C. Kyrpides, L.L. Blackall, K.D. McMahon, P. Hugenholtz, Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities, Nat. Biotechnol. 24 (10) (2006) 1263–1269.
[5] A. Oehmen, P.C. Lemos, G. Carvalho, Z. Yuan, J. Keller, L.L. Blackall, M.A.M. Reis, Advances in enhanced biological phosphorus removal: from micro to macro scale, Water Res. 41 (11) (2007) 2271–2300.
[6] G. Qiu, R. Zuniga-Montanez, Y. Law, S.S. Thi, T.Q.N. Nguyen, K. Eganathan, X. Liu, P.H. Nielsen, R.B.H. Williams, S. Wuertz, Polyphosphate-accumulating organisms in full-scale tropical wastewater treatment plants use diverse carbon sources, Water Res. 149 (2019) 496–510.
[7] G. Qiu, X. Liu, N.M.M.T. Saw, Y. Law, R. Zuniga-Montanez, S.S. Thi, T.Q. Ngoc Nguyen, P.H. Nielsen, R.B.H. Williams, S. Wuertz, Metabolic traits of Candidatus Accumulibacter clade IIF strain SCELSE-1 using amino acids as carbon sources for enhanced biological phosphorus removal, Environ. Sci. Technol. 54 (4) (2020) 2448–2458.
[8] L. Chen, H. Chen, Z. Hu, Y. Tian, C. Wang, P. Xie, X. Deng, Y. Zhang, X. Tang, X. Lin, B. Li, C. Wei, G. Qiu, Carbon uptake bioenergetics of PAOs and GAOs in full-scale enhanced biological phosphorus removal systems, Water Res. 216 (2022) 118258.
[9] R. Diaz, B. Mackey, S. Chadalavada, J. kainthola, P. Heck, R. Goel, Enhanced Bio-P removal: past, present, and future – a comprehensive review, Chemosphere 309 (2022) 136518.
[10] C. Zhang, A. Guisasola, J.A. Baeza, A review on the integration of mainstream P-recovery strategies with enhanced biological phosphorus removal, Water Res. 212 (2022) 118102.
[11] S. He, K.D. McMahon, Microbiology of 'Candidatus Accumulibacter' in activated sludge, Microb. Biotechnol. 4 (5) (2011) 603–619.
[12] P.H. Nielsen, S.J. McIlroy, M. Albertsen, M. Nierychlo, Re-evaluating the microbiology of the enhanced biological phosphorus removal process, Curr. Opin. Biotechnol. 57 (2019) 111–118.
[13] A.G. Dorofeev, Y.A. Nikolaev, A.V. Mardanov, N.V. Pimenov, Role of phosphate-accumulating bacteria in biological phosphorus removal from wastewater, Appl. Biochem. Microbiol. 56 (1) (2020) 1–14.
[14] C. Zhang, X. Chen, M. Han, X. Li, H. Chang, N. Ren, S.-H. Ho, Revealing the role of microalgae-bacteria niche for boosting wastewater treatment and energy reclamation in response to temperature, Environmental Science and Ecotechnology 14 (2023) 100230.
[15] R.J. Seviour, T. Mino, M. Onuki, The microbiology of biological phosphorus removal in activated sludge systems, FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 27 (1) (2003) 99–127.
[16] Y. Mao, D.W. Graham, H. Tamaki, T. Zhang, Dominant and novel clades of Candidatus Accumulibacter phosphatis in 18 globally distributed full-scale wastewater treatment plants, Sci. Rep. 5 (1) (2015) 11857.
[17] S. Roy, Q. Guanglei, R. Zuniga-Montanez, R.B.H. Williams, S. Wuertz, Recent advances in understanding the ecophysiology of enhanced biological phosphorus removal, Curr. Opin. Biotechnol. 67 (2021) 166–174.
[18] F. Petriglieri, C.M. Singleton, Z. Kondrotaite, M.K.D. Dueholm, E.A. McDaniel, K.D. McMahon, P.H. Nielsen, Reevaluation of the phylogenetic diversity and global distribution of the genus Candidatus Accumulibacter, mSystems 7 (3) (2022) e00016-00022.
[19] A. Oehmen, R.J. Zeng, Z. Yuan, J. Keller, Anaerobic metabolism of propionate by polyphosphate-accumulating organisms in enhanced biological phosphorus removal systems, Biotechnol. Bioeng. 91 (1) (2005) 43–53.
[20] S. Kolakovic, E.B. Freitas, M.A.M. Reis, G. Carvalho, A. Oehmen, Accumulibacter diversity at the sub-clade level impacts enhanced biological phosphorus removal performance, Water Res. 199 (2021) 117210.
[21] W. Zhao, X. Bi, Y. Peng, M. Bai, Research advances of the phosphorus-accumulating organisms of Candidatus Accumulibacter, Dechloromonas and Tetrasphaera: metabolic mechanisms, applications and influencing factors, Chemosphere 307 (2022) 135675.
[22] T. Páez-Watson, M.C.M. van Loosdrecht, S.A. Wahl, Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: application to polyphosphate accumulating organisms, Water Res. 228 (2023) 119365.
[23] I. Bessarab, A.M. Maszenan, M.A.S. Haryono, K. Arumugam, N.M.M.T. Saw, R.J. Seviour, R.B.H. Williams, Comparative genomics of members of the genus Defluviicoccus with insights into their ecophysiological importance, Front. Microbiol. 13 (2022) 834906.
[24] A.M. Maszenan, I. Bessarab, R.B.H. Williams, S. Petrovski, R.J. Seviour, The phylogeny, ecology and ecophysiology of the glycogen accumulating organism (GAO) Defluviicoccus in wastewater treatment plants, Water Res. 221 (2022) 118729.
[25] M.M. Turcotte, M.S.C. Corrin, M.T.J. Johnson, Adaptive evolution in ecological communities, PLoS Biol. 10 (5) (2012) e1001332.
[26] B.O. Oyserman, D.R. Noguera, T.G. del Rio, S.G. Tringe, K.D. McMahon, Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis, ISME J. 10 (4) (2016) 810–822.
[27] H. Tettelin, V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S.V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. DeBoy, T.M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S.C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J.B. O'Connor, S. Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M.R. Wessels, R. Rappuoli, C.M. Fraser, Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome", Proc. Natl. Acad. Sci. USA 102 (39) (2005) 13950–13955.
[28] J.-M. Song, Z. Guan, J. Hu, C. Guo, Z. Yang, S. Wang, D. Liu, B. Wang, S. Lu, R. Zhou, W.-Z. Xie, Y. Cheng, Y. Zhang, K. Liu, Q.-Y. Yang, L.-L. Chen, L. Guo, Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus, Nat. Plants 6 (1) (2020) 34–45.
[29] D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome, Curr. Opin. Genet. Dev. 15 (6) (2005) 589–594.
[30] R. Della Coletta, Y. Qiu, S. Ou, M.B. Hufford, C.N. Hirsch, How the pan-genome is changing crop genomics and improvement, Genome Biol. 22 (1) (2021) 3.
[31] S.K. Aggarwal, A. Singh, M. Choudhary, A. Kumar, S. Rakshit, P. Kumar, A. Bohra, R.K. Varshney, Pangenomics in microbial and crop research: progress, applications, and perspectives, Genes 13 (4) (2022) 598.
[32] A.A. Golicz, J. Batley, D. Edwards, Towards plant pangenomics, Plant Biotechnol. J. 14 (4) (2016) 1099–1105.
[33] J.J. Flowers, S. He, S. Malfatti, T.G. del Rio, S.G. Tringe, P. Hugenholtz, K.D. McMahon, Comparative genomics of two 'Candidatus Accumulibacter' clades performing biological phosphorus removal, ISME J. 7 (12) (2013) 2301–2314.
[34] P.Y. Camejo, B.O. Oyserman, K.D. McMahon, D.R. Noguera, Integrated omic analyses provide evidence that a "Candidatus Accumulibacter phosphatis" strain performs denitrification under microaerobic conditions, mSystems 4 (1) (2019) e00193-00118.
[35] N.M. El-Sayed, P.J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E.A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D.C. Bartholomeu, B.J. Haas, A.-N. Tran, J.R. Wortman, U.C.M. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J.M. Carlton, G.C. Cerqueira, T. Creasy, A.L. Delcher, A. Djikeng, T.M. Embley, C. Hauser, A.C. Ivens, S.K. Kummerfeld, J.B. Pereira-Leal, D. Nilsson, J. Peterson, S.L. Salzberg, J. Shallom, J.C. Silva, J. Sundaram, S. Westenberger, O. White, S.E. Melville, J.E. Donelson, B. Andersson, K.D. Stuart, N. Hall, Comparative genomics of trypanosomatid parasitic protozoa, Science 309 (5733) (2005) 404–409.
[36] B. Fernández-Gómez, M. Richter, M. Schüler, J. Pinhassi, S.G. Acinas, J.M. González, C. Pedrós-Alió, Ecology of marine Bacteroidetes: a comparative genomics approach, ISME J. 7 (5) (2013) 1026–1037.
[37] A. Coghlan, R. Tyagi, J.A. Cotton, N. Holroyd, B.A. Rosa, I.J. Tsai, D.R. Laetsch, R.N. Beech, T.A. Day, K. Hallsworth-Pepin, H.-M. Ke, T.-H. Kuo, T.J. Lee, J. Martin, R.M. Maizels, P. Mutowo, P. Ozersky, J. Parkinson, A.J. Reid, N.D. Rawlings, D.M. Ribeiro, L.S. Swapna, E. Stanley, D.W. Taylor, N.J. Wheeler, M. Zamanian, X. Zhang, F. Allan, J.E. Allen, K. Asano, S.A. Babayan, G. Bah, H. Beasley, H.M. Bennett, S.A. Bisset, E. Castillo, J. Cook, P.J. Cooper, T. Cruz-Bustos, C. Cuéllar, E. Devaney, S.R. Doyle, M.L. Eberhard, A. Emery, K.S. Eom, J.S. Gilleard, D. Gordon, Y. Harcus, B. Harsha, J.M. Hawdon, D.E. Hill, J. Hodgkinson, P. Horák, K.L. Howe, T. Huckvale, M. Kalbe, G. Kaur, T. Kikuchi, G. Koutsovoulos, S. Kumar, A.R. Leach, J. Lomax, B. Makepeace, J.B. Matthews, A. Muro, N.M. O'Boyle, P.D. Olson, A. Osuna, F. Partono, K. Pfarr, G. Rinaldi, P. Foronda, D. Rollinson, M.G. Samblas, H. Sato, M. Schnyder, T. Scholz, M. Shafie, V.N. Tanya, R. Toledo, A. Tracey, J.F. Urban, L.-C. Wang, D. Zarlenga, M.L. Blaxter, M. Mitreva, M. Berriman, C. International Helminth Genomes, Comparative genomics of the major parasitic worms, Nat. Genet. 51 (1) (2019) 163–174.
[38] I. Kjærbølling, T. Vesth, J.C. Frisvad, J.L. Nybo, S. Theobald, S. Kildgaard, T.I. Petersen, A. Kuo, A. Sato, E.K. Lyhne, M.E. Kogle, A. Wiebenga, R.S. Kun, R.J.M. Lubbers, M.R. Mäkelä, K. Barry, M. Chovatia, A. Clum, C. Daum, S. Haridas, G. He, K. LaButti, A. Lipzen, S. Mondo, J. Pangilinan, R. Riley, A. Salamov, B.A. Simmons, J.K. Magnuson, B. Henrissat, U.H. Mortensen, T.O. Larsen, R.P. de Vries, I.V. Grigoriev, M. Machida, S.E. Baker, M.R. Andersen, A comparative genomics study of 23 Aspergillus species from section Flavi, Nat. Commun. 11 (1) (2020) 1106.
[39] S. Feng, J. Stiller, Y. Deng, J. Armstrong, Q. Fang, A.H. Reeve, D. Xie, G. Chen, C. Guo, B.C. Faircloth, B. Petersen, Z. Wang, Q. Zhou, M. Diekhans, W. Chen, S. Andreu-Sánchez, A. Margaryan, J.T. Howard, C. Parent, G. Pacheco, M.-H.S. Sinding, L. Puetz, E. Cavill, A.M. Ribeiro, L. Eckhart, J. Fjeldså, P.A. Hosner, R.T. Brumfield, L. Christidis, M.F. Bertelsen, T. Sicheritz-Ponten, D.T. Tietze, B.C. Robertson, G. Song, G. Borgia, S. Claramunt, I.J. Lovette, S.J. Cowen, P. Njoroge, J.P. Dumbacher, O.A. Ryder, J. Fuchs, M. Bunce, D.W. Burt, J. Cracraft, G. Meng, S.J. Hackett, P.G. Ryan, K.A. Jønsson, I.G. Jamieson, R.R. da Fonseca, E.L. Braun, P. Houde, S. Mirarab, A. Suh, B. Hansson, S. Ponnikas, H. Sigeman, M. Stervander, P.B. Frandsen, H. van der Zwan, R. van der Sluis,

C. Visser, C.N. Balakrishnan, A.G. Clark, J.W. Fitzpatrick, R. Bowman, N. Chen, A. Cloutier, T.B. Sackton, S.V. Edwards, D.J. Foote, S.B. Shakya, F.H. Sheldon, A. Vignal, A.E.R. Soares, B. Shapiro, J. González-Solís, J. Ferrer-Obiol, J. Rozas, M. Riutort, A. Tigano, V. Friesen, L. Dalén, A.O. Urrutia, T. Székely, Y. Liu, M.G. Campana, A. Corvelo, R.C. Fleischer, K.M. Rutherford, N.J. Gemmell, N. Dussex, H. Mouritsen, N. Thiele, K. Delmore, M. Liedvogel, A. Franke, M.P. Hoeppner, O. Krone, A.M. Fudickar, B. Milá, E.D. Ketterson, A.E. Fidler, G. Friis, Á.M. Parody-Merino, P.F. Battley, M.P. Cox, N.C.B. Lima, F. Prosdocimi, T.L. Parchman, B.A. Schlinger, B.A. Loiselle, J.G. Blake, H.C. Lim, L.B. Day, M.J. Fuxjager, M.W. Baldwin, M.J. Braun, M. Wirthlin, R.B. Dikow, T.B. Ryder, G. Camenisch, L.F. Keller, J.M. DaCosta, M.E. Hauber, M.I.M. Louder, C.C. Witt, J.A. McGuire, J. Mudge, L.C. Megna, M.D. Carling, B. Wang, S.A. Taylor, G. Del-Rio, A. Aleixo, A.T.R. Vasconcelos, C.V. Mello, J.T. Weir, D. Haussler, Q. Li, H. Yang, J. Wang, F. Lei, C. Rahbek, M.T.P. Gilbert, G.R. Graves, E.D. Jarvis, B. Paten, G. Zhang, Dense sampling of bird diversity increases power of comparative genomics, Nature 587 (7833) (2020) 252–257.

[40] Z. Zhang, Y. Guo, F. Yang, J. Li, Pan-Genome analysis reveals functional divergences in gut-restricted Gilliamella and Snodgrassella, Bioengineering 9 (10) (2022) 544.

[41] E. Fernandez-Fueyo, F.J. Ruiz-Dueñas, P. Ferreira, D. Floudas, D.S. Hibbett, P. Canessa, L.F. Larrondo, T.Y. James, D. Seelenfreund, S. Lobos, R. Polanco, M. Tello, Y. Honda, T. Watanabe, T. Watanabe, J.S. Ryu, C.P. Kubicek, M. Schmoll, J. Gaskell, K.E. Hammel, St John FJ, A. Vanden Wymelenberg, G. Sabat, S. Splinter BonDurant, K. Syed, J.S. Yadav, H. Doddapaneni, V. Subramanian, J.L. Lavín, J.A. Oguiza, G. Perez, A.G. Pisabarro, L. Ramirez, F. Santoyo, E. Master, P.M. Coutinho, B. Henrissat, V. Lombard, J.K. Magnuson, U. Kües, C. Hori, K. Igarashi, M. Samejima, B.W. Held, K.W. Barry, K.M. LaButti, A. Lapidus, E.A. Lindquist, S.M. Lucas, R. Riley, A.A. Salamov, D. Hoffmeister, D. Schwenk, Y. Hadar, O. Yarden, R.P. de Vries, A. Wiebenga, J. Stenlid, D. Eastwood, I.V. Grigoriev, R.M. Berka, R.A. Blanchette, P. Kersten, A.T. Martinez, R. Vicuna, D. Cullen, Comparative genomics of Ceriporiopsis subvermispora and Phanerochaete chrysosporium provide insight into selective ligninolysis, Proc. Natl. Acad. Sci. USA 109 (14) (2012) 5458–5463.

[42] B.O. Oyserman, F. Moya, C.E. Lawson, A.L. Garcia, M. Vogt, M. Heffernen, D.R. Noguera, K.D. McMahon, Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria, ISME J. 10 (12) (2016) 2931–2945.

[43] K. Arumugam, C. Bağcı, I. Bessarab, S. Beier, B. Buchfink, A. Górska, G. Qiu, D.H. Huson, R.B.H. Williams, Annotated bacterial chromosomes from frameshift-corrected long-read metagenomic data, Microbiome 7 (1) (2019) 61.

[44] F.J. Rubio-Rincón, D.G. Weissbrodt, C.M. Lopez-Vazquez, L. Welles, B. Abbas, M. Albertsen, P.H. Nielsen, M.C.M. van Loosdrecht, D. Brdjanovic, "Candidatus Accumulibacter delftensis": a clade IC novel polyphosphate-accumulating organism without denitrifying activity on nitrate, Water Res. 161 (2019) 136–151.

[45] E.A. McDaniel, F. Moya-Flores, N. Keene Beach, P.Y. Camejo, B.O. Oyserman, M. Kizaric, E.H. Khor, D.R. Noguera, K.D. McMahon, Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics, mSystems 6 (4) (2021) e0047421.

[46] C.M. Singleton, F. Petriglieri, J.M. Kristensen, R.H. Kirkegaard, T.Y. Michaelsen, M.H. Andersen, Z. Kondrotaite, S.M. Karst, M.S. Dueholm, P.H. Nielsen, M. Albertsen, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing, Nat. Commun. 12 (1) (2021) 2009.

[47] V.N. Srinivasan, G. Li, D. Wang, N.B. Tooker, Z. Dai, A. Onnis-Hayden, C. Bott, P. Dombrowski, P. Schauer, A. Pinto, A.Z. Gu, Oligotyping and metagenomics reveal distinct Candidatus Accumulibacter communities in side-stream versus conventional full-scale enhanced biological phosphorus removal (EBPR) systems, Water Res. 206 (2021) 117725.

[48] Y. Tian, H. Chen, L. Chen, X. Deng, Z. Hu, C. Wang, C. Wei, G. Qiu, S. Wuertz, Glycine adversely affects enhanced biological phosphorus removal, Water Res. 209 (2022) 117894.

[49] X. Deng, J. Yuan, L. Chen, H. Chen, C. Wei, P.H. Nielsen, S. Wuertz, G. Qiu, CRISPR-Cas phage defense systems and prophages in Candidatus Accumulibacter, Water Res. 235 (2023) 119906.

[50] M. Albertsen, S.J. McIlroy, M. Stokholm-Bjerregaard, S.M. Karst, P.H. Nielsen, "Candidatus Propionivibrio aalborgensis": a novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants, Front. Microbiol. 7 (2016) 1033.

[51] F. Petriglieri, C. Singleton, M. Peces, J.F. Petersen, M. Nierychlo, P.H. Nielsen, "Candidatus Dechloromonas phosphoritropha" and "Ca. D. phosphorivorans", novel polyphosphate accumulating organisms abundant in wastewater treatment systems, ISME J. 15 (12) (2021) 3605–3614.

[52] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (7) (2015) 1043–1055.

[53] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (1) (2019) 238.

[54] A.J. Enright, S. Van Dongen, C.A. Ouzounis, An efficient algorithm for large-scale detection of protein families, Nucleic Acids Res. 30 (7) (2002) 1575–1584.

[55] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780.

[56] J. Castresana, Selection of conserved blocks from multiple alignments for their sse in phylogenetic analysis, Mol. Biol. Evol. 17 (4) (2000) 540–552.

[57] W. Shen, S. Le, Y. Li, F. Hu, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, PLoS One 11 (10) (2016) e0163962.

[58] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (5) (2020) 1530–1534.

[59] I. Letunic, P. Bork, Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, Nucleic Acids Res. 49 (W1) (2021) W293–W296.

[60] A.-N. Zhang, Y. Mao, Y. Wang, T. Zhang, Mining traits for the enrichment and isolation of not-yet-cultured populations, Microbiome 7 (1) (2019) 96.

[61] M. Csűös, Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood, Bioinformatics 26 (15) (2010) 1910–1912.

[62] C. Pál, B. Papp, M.J. Lercher, Adaptive evolution of bacterial metabolic networks by horizontal gene transfer, Nat. Genet. 37 (12) (2005) 1372–1375.

[63] K. Zaremba-Niedzwiedzka, J. Viklund, W. Zhao, J. Ast, A. Sczyrba, T. Woyke, K. McMahon, S. Bertilsson, R. Stepanauskas, S.G. Andersson, Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade, Genome Biol. 14 (11) (2013) 1–14.

[64] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information, knowledge and principle: back to metabolism in KEGG, Nucleic Acids Res. 42 (D1) (2013) D199–D205.

[65] M. Ravenhall, N. Škunca, F. Lassalle, C. Dessimoz, Inferring horizontal gene transfer, PLoS Comput. Biol. 11 (5) (2015) e1004095.

[66] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 35 (suppl_1) (2006) D61–D65.

[67] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.

[68] E. Kopylova, L. Noé, H. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, Bioinformatics 28 (24) (2012) 3211–3217.

[69] Bushnell B: BBMap: a fast, accurate, splice-aware aligner. In: Conference: 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17-20, 2014; United States. DE-AC02-05CH11231 2016-04-08: 2014: Medium: ED.

[70] L. Kumar, E.F. M, Mfuzz: a software package for soft clustering of microarray data, Bioinformation 2 (1) (2007) 5–7.

[71] Y. Shang, X. Wang, Z. Chen, Z. Lyu, Z. Lin, J. Zheng, Y. Wu, Q. Deng, Z. Yu, Y. Zhang, D. Qu, Staphylococcus aureus PhoU homologs regulate persister formation and virulence, Front. Microbiol. 11 (2020) 865.

[72] L. Chen, G. Wei, Y. Zhang, K. Wang, C. Wang, X. Deng, Y. Li, X. Xie, J. Chen, F. Huang, H. Chen, B. Zhang, C. Wei, G. Qiu, Candidatus Accumulibacter use fermentation products for enhanced biological phosphorus removal, Water Res. 246 (2023) 120713.

[73] Y. Zhou, M. Pijuan, R.J. Zeng, Z. Yuan, Involvement of the TCA cycle in the anaerobic metabolism of polyphosphate accumulating organisms (PAOs), Water Res. 43 (5) (2009) 1330–1340.

[74] E.J. Capra, M.T. Laub, Evolution of two-component signal transduction systems, Annu. Rev. Microbiol. 66 (2012) 325–347.

[75] E. Alm, K. Huang, A. Arkin, The evolution of two-component systems in bacteria reveals different strategies for niche adaptation, PLoS Comput. Biol. 2 (11) (2006) e143.

[76] L.E. Ulrich, I.B. Zhulin, The MiST2 database: a comprehensive genomics resource on microbial signal transduction, Nucleic Acids Res. 38 (Database issue) (2010). D401-407.

[77] S. Elsen, L.R. Swem, D.L. Swem, C.E. Bauer, RegB/RegA, a highly conserved redox-responding global two-component regulatory system, Microbiol. Mol. Biol. Rev. 68 (2) (2004) 263–279.

[78] S. Elsen, W. Dischert, A. Colbeau, C.E. Bauer, Expression of uptake hydrogenase and molybdenum nitrogenase in Rhodobacter capsulatus is coregulated by the RegB-RegA two-component regulatory system, J. Bacteriol. 182 (10) (2000) 2831–2837.

[79] J.M. Dubbs, T.H. Bird, C.E. Bauer, F.R. Tabita, Interaction of CbbR and RegA* transcription regulators with the Rhodobacter sphaeroides cbb promoter-operator region *, J. Biol. Chem. 275 (25) (2000) 19224–19230.

[80] F. Santos-Beneit, The Pho regulon: a huge regulatory network in bacteria, Front. Microbiol. 6 (2015) 402.

[81] S. Choi, G. Jeong, E. Choi, E.-J. Lee, A dual regulatory role of the PhoU protein in Salmonella Typhimurium, mBio 13 (3) (2022) e00811-00822.

[82] V. Monzon, T. Paysan-Lafosse, V. Wood, A. Bateman, Reciprocal best structure hits: using AlphaFold models to discover distant homologues, Bioinformatics Advances 2 (1) (2022) vbac072.

[83] G.C. diCenzo, H. Sharthiya, A. Nanda, M. Zamani, T.M. Finan, PhoU allows rapid adaptation to high phosphate concentrations by modulating PstSCAB transport rate in Sinorhizobium meliloti, J. Bacteriol. 199 (18) (2017).

[84] Y. Li, Y. Zhang, PhoU is a persistence switch involved in persister formation and tolerance to multiple antibiotics and stresses in Escherichia coli, Antimicrob. Agents Chemother. 51 (6) (2007) 2092–2099.

[85] F.A. Herbst, M.S. Dueholm, R. Wimmer, P.H. Nielsen, The proteome of Tetrasphaera elongata is adapted to changing conditions in wastewater treatment plants, Proteomes 7 (2) (2019).

[86] G.R. Willsky, M.H. Malamy, Characterization of two genetically separable inorganic phosphate transport systems in Escherichia coli, J. Bacteriol. 144 (1)

(1980) 356–365.
[87] J.F. Martín, P. Liras, Molecular mechanisms of phosphate sensing, transport and signalling in Streptomyces and related Actinobacteria, Int. J. Mol. Sci. 22 (3) (2021).
[88] S. Gebhard, S.L. Tran, G.M. Cook, The Phn system of Mycobacterium smegmatis: a second high-affinity ABC-transporter for phosphate, Microbiology 152 (11) (2006) 3453–3465.
[89] A.M. Saunders, A.N. Mabbett, A.G. McEwan, L.L. Blackall, Proton motive force generation from stored polymers for the uptake of acetate under anaerobic conditions, FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 274 (2) (2007) 245–251.
[90] Z. Wang, F. Guo, Y. Mao, Y. Xia, T. Zhang, Metabolic characteristics of a glycogen-accumulating organism in Defluviicoccus cluster II revealed by comparative genomics, Microb. Ecol. 68 (4) (2014) 716–728.
[91] S.J. McIlroy, M. Albertsen, E.K. Andresen, A.M. Saunders, R. Kristiansen, M. Stokholm-Bjerregaard, K.L. Nielsen, P.H. Nielsen, 'Candidatus Competibacter'-lineage genomes retrieved from metagenomes reveal functional metabolic diversity, ISME J. 8 (3) (2014) 613–624.